Chapter 9: Assumption-based Reasoning

- Lecture 1 Assumption-based reasoning framework.
- Lecture 2 Default reasoning, the multiple-extension problem, skeptical reasoning.
- Lecture 3 Abduction, abductibe diagnosis
- Lecture 4 Combining Evidential and Causal Reasoning
- Lecture 5 Algorithms

Assumption-based Reasoning

Often we want our agents to make assumptions rather than doing deduction from their knowledge. For example:

- In default reasoning the delivery robot may want to assume Mary is in her office, even if it isn't always true.
- In diagnosis you hypothesize what could be wrong with a system to produce the observed symptoms.
- In design you hypothesize components that provably fulfill some design goals and are feasible.

Design and Recognition

Two different tasks use assumption-based reasoning:

- **Design** The aim is to design an artifact or plan. The designer can select whichever design they like that satisfies the design criteria.
- **Recognition** The aim is to find out what is true based on observations. If there are a number of possibilities, the recognizer can't select the one they like best. The underlying reality is fixed; the aim is to find out what it is.

Compare: Recognizing a disease with designing a treatment. Designing a meeting time with determining when it is.

The Assumption-based Framework

The assumption-based framework is defined in terms of two sets of formulae:

- F is a set of closed formula called the facts.
  These are formulae that are given as true in the world.
  We assume F are Horn clauses.
- H is a set of formulae called the possible hypotheses or assumables. Ground instance of the possible hypotheses can be assumed if consistent.

Making Assumptions

- A **scenario** of $\langle F, H \rangle$ is a set D of ground instances of elements of H such that $F \cup D$ is satisfiable.

- An **explanation** of g from $\langle F, H \rangle$ is a scenario that, together with F, implies g.

  D is an explanation of g if $F \cup D \models g$ and $F \cup D \not\models \mathit{false}$.

  A **minimal explanation** is an explanation such that no strict subset is also an explanation.

- An **extension** of $\langle F, H \rangle$ is the set of logical consequences of F and a maximal scenario of $\langle F, H \rangle$.

Example

$a \leftarrow b \wedge c.$

$b \leftarrow e.$

$b \leftarrow h.$

$c \leftarrow g.$

$c \leftarrow f.$

$d \leftarrow g.$

$false \leftarrow e \wedge d.$

$f \leftarrow h \wedge m.$

assumable $e, h, g, m, n.$

- $\{e, m, n\}$ is a scenario.
- $\{e, g, m\}$ is not a scenario.
- $\{h, m\}$ is an explanation for a.
- $\{e, h, m\}$ is an explanation for a.
- $\{e, h, m, n\}$ is a maximal scenario.
- $\{h, g, m, n\}$ is a maximal scenario.

Default Reasoning and Abduction

There are two strategies for using the assumption-based framework:

- **Default reasoning** Where the truth of g is unknown and is to be determined.

  An explanation for g corresponds to an **argument** for g.

- **Abduction** Where g is given, and we are interested in explaining it. g could be an observation in a recognition task or a design goal in a design task.

Default Reasoning

- When giving information, you don't want to enumerate all of the exceptions, even if you could think of them all.
- In default reasoning, you specify general knowledge and modularly add exceptions. The general knowledge is used for cases you don't know are exceptional.
- Classical logic is monotonic: If g logically follows from A, it also follows from any superset of A.
- Default reasoning is nonmonotonic: When you add that something is exceptional, you can't conclude what you could before.

Defaults as Assumptions

Default reasoning can be modeled using

- H is normality assumptions
- F states what follows from the assumptions

An explanation of g gives an argument for g.

Default Example

A reader of newsgroups may have a default:
"Articles about AI are generally interesting".

$$H = \{int\_ai(X)\},$$

where $int\_ai(X)$ means X is interesting if it is about AI.

With facts:

$$interesting(X) \leftarrow about\_ai(X) \wedge int\_ai(X).$$

$$about\_ai(art\_23).$$

$\{int\_ai(art\_23)\}$ is an explanation for $interesting(art\_23)$.

Default Example, Continued

We can have exceptions to defaults:

$$false \leftarrow interesting(X) \wedge uninteresting(X).$$

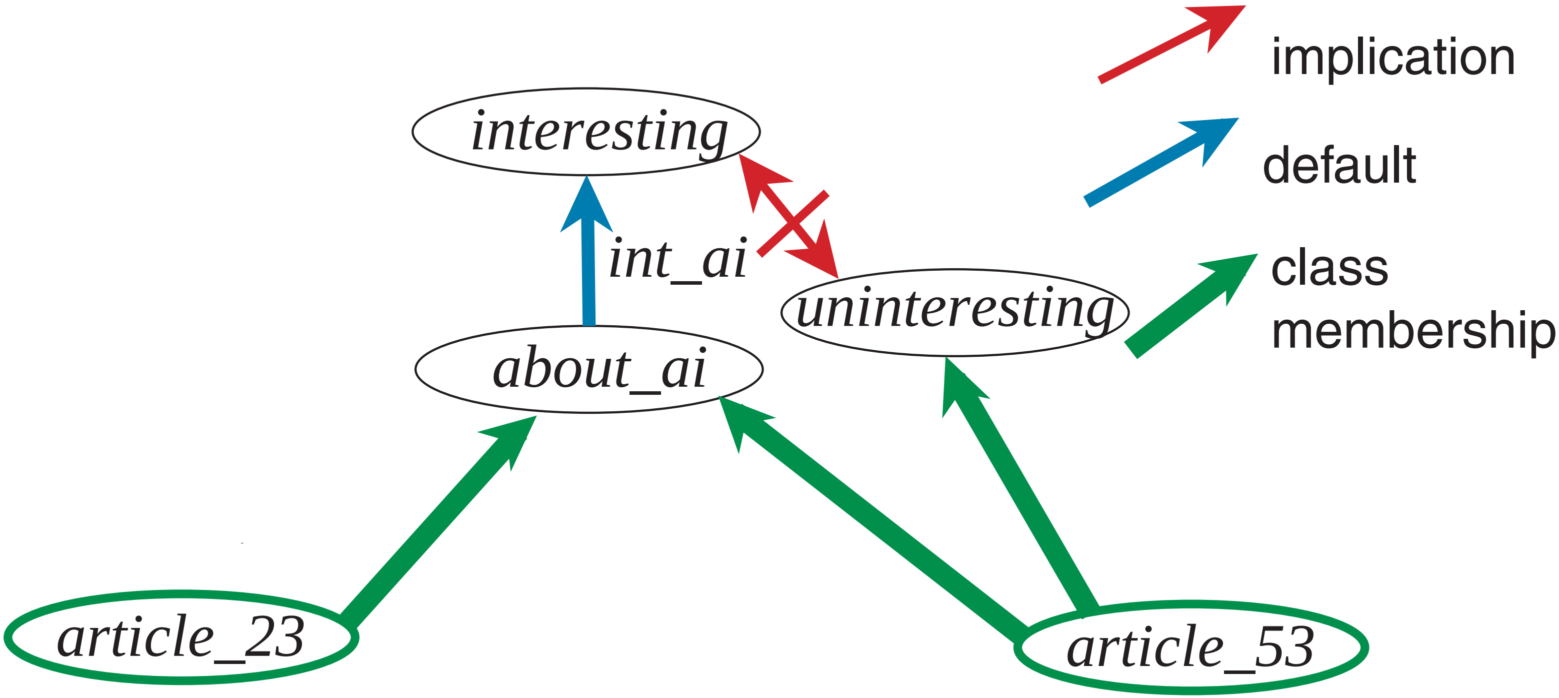
Suppose article 53 is about AI but is uninteresting:

$$about\_ai(art\_53).$$

$$uninteresting(art\_53).$$

We cannot explain $interesting(art\_53)$ even though everything we know about art_23 you also know about art_53.

Exceptions to defaults



Exceptions to Defaults

"Articles about formal logic are about AI."

"Articles about formal logic are uninteresting."

"Articles about machine learning are about AI."

$$about\_ai(X) \leftarrow about\_fl(X).$$

$$uninteresting(X) \leftarrow about\_fl(X).$$

$$about\_ai(X) \leftarrow about\_ml(X).$$

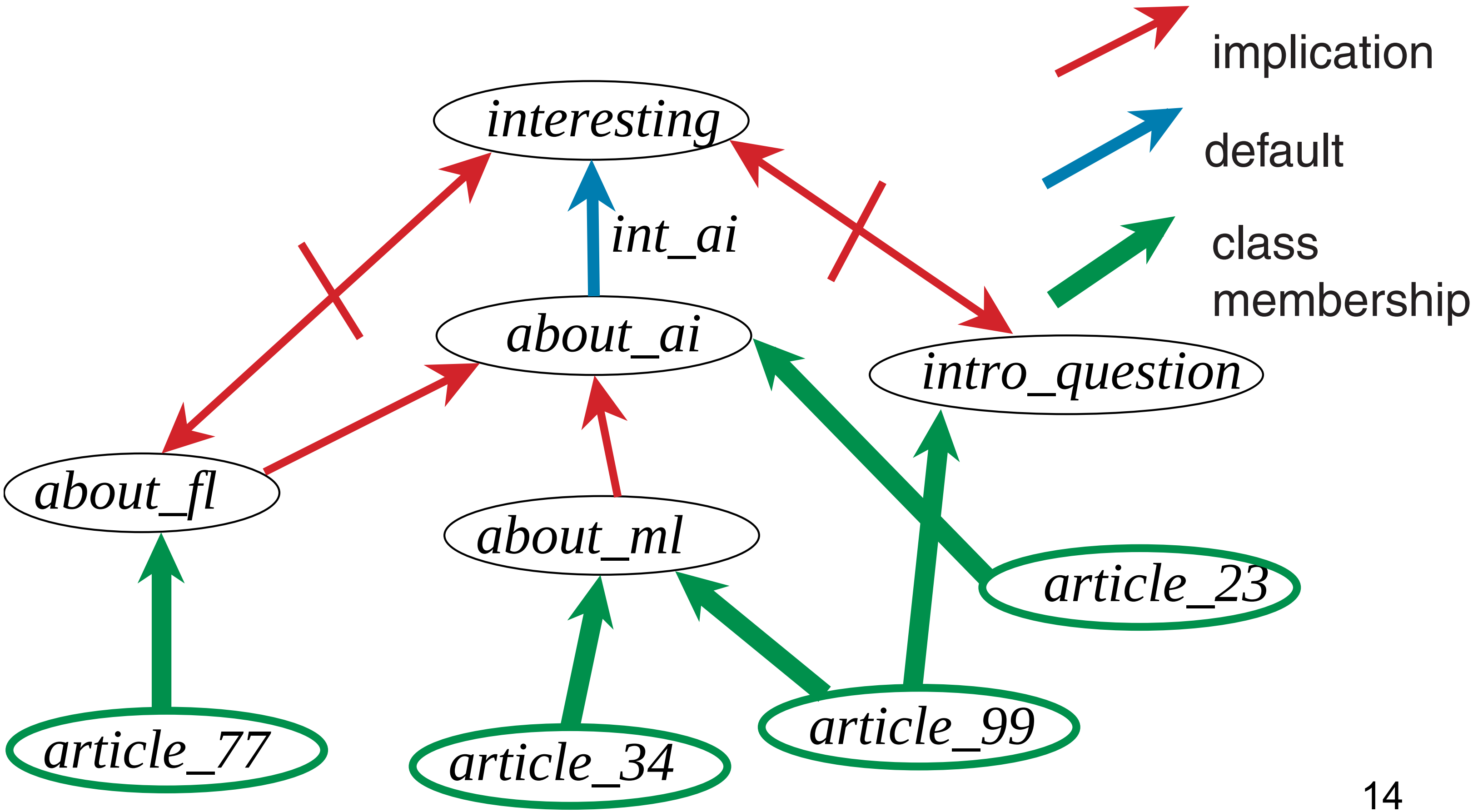
$$about\_fl(art\_77).$$

$$about\_ml(art\_34).$$

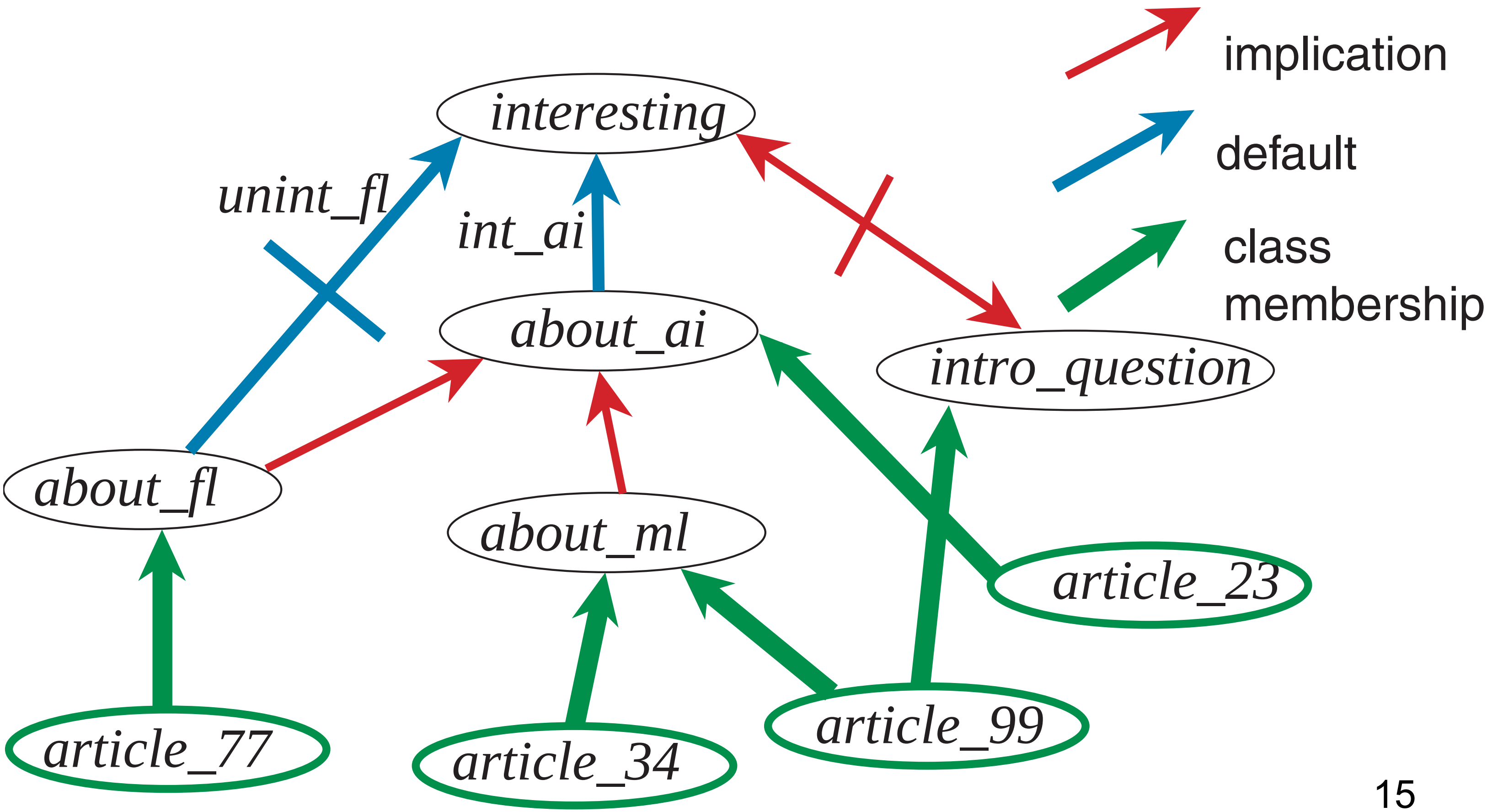
You can't explain $interesting(art\_77)$.

You can explain $interesting(art\_34)$.

Exceptions to Defaults



Formal logic is uninteresting by default



Contradictory Explanations

Suppose formal logic articles aren't interesting *by default*:

$$H = \{unint\_fl(X), int\_ai(X)\}$$

The corresponding facts are:

$$interesting(X) \leftarrow about\_ai(X) \wedge int\_ai(X).$$

$$about\_ai(X) \leftarrow about\_fl(X).$$

$$uninteresting(X) \leftarrow about\_fl(X) \wedge unint\_fl(X).$$

$$about\_fl(art\_77).$$

$uninteresting(art\_77)$ has explanation $\{unint\_fl(art\_77)\}$.

$interesting(art\_77)$ has explanation $\{int\_ai(art\_77)\}$.

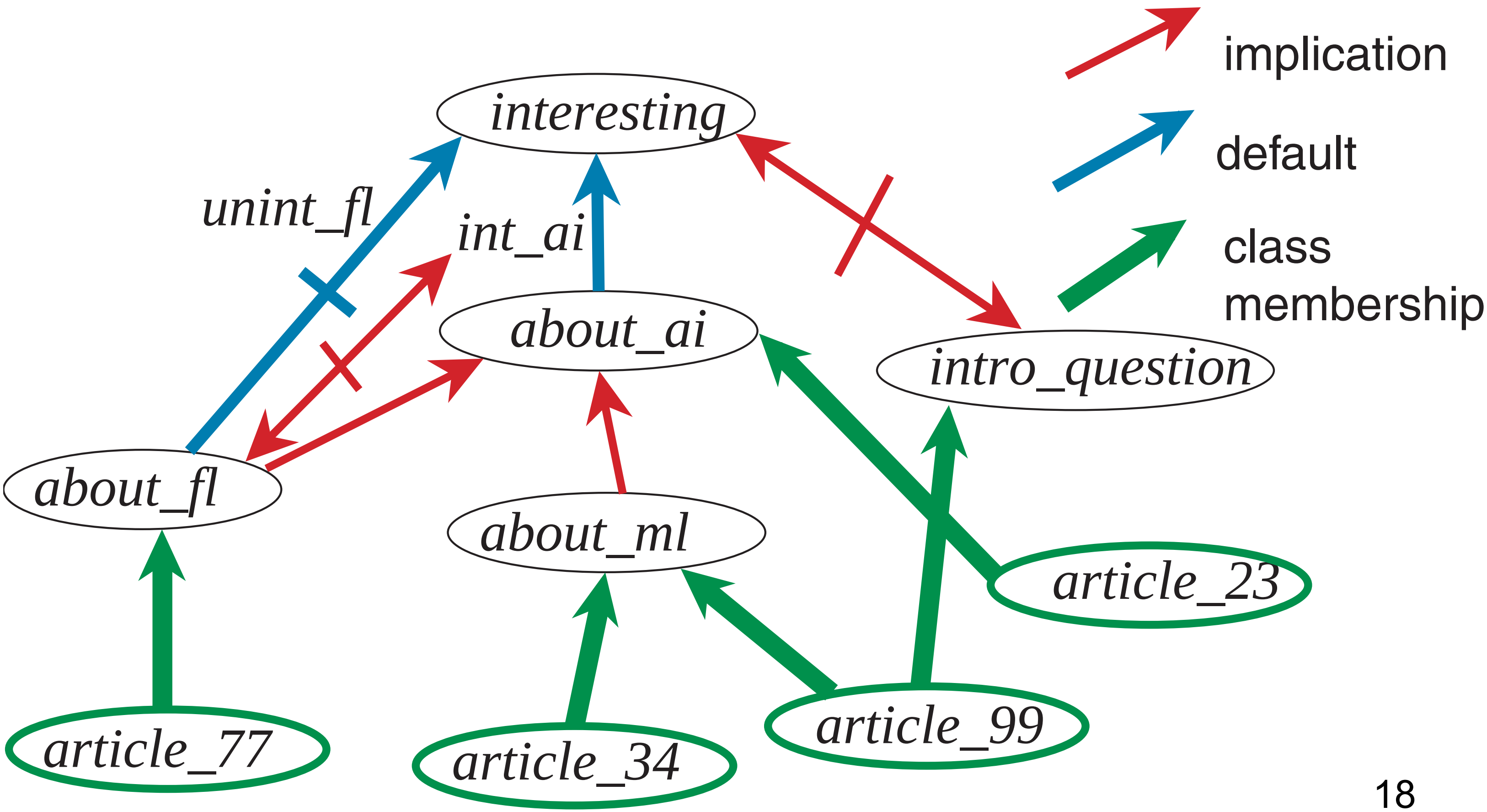
Overriding Assumptions

- Because *art_77* is about formal logic, the argument "*art_77* is interesting because it is about AI" shouldn't be applicable.
- This is an instance of preference for more specific defaults.
- Arguments that articles about formal logic are interesting because they are about AI can be defeated by adding:

$$false \leftarrow about\_fl(X) \wedge int\_ai(X).$$

  This is known as a cancellation rule.
- You can no longer explain *interesting*(*art_77*).

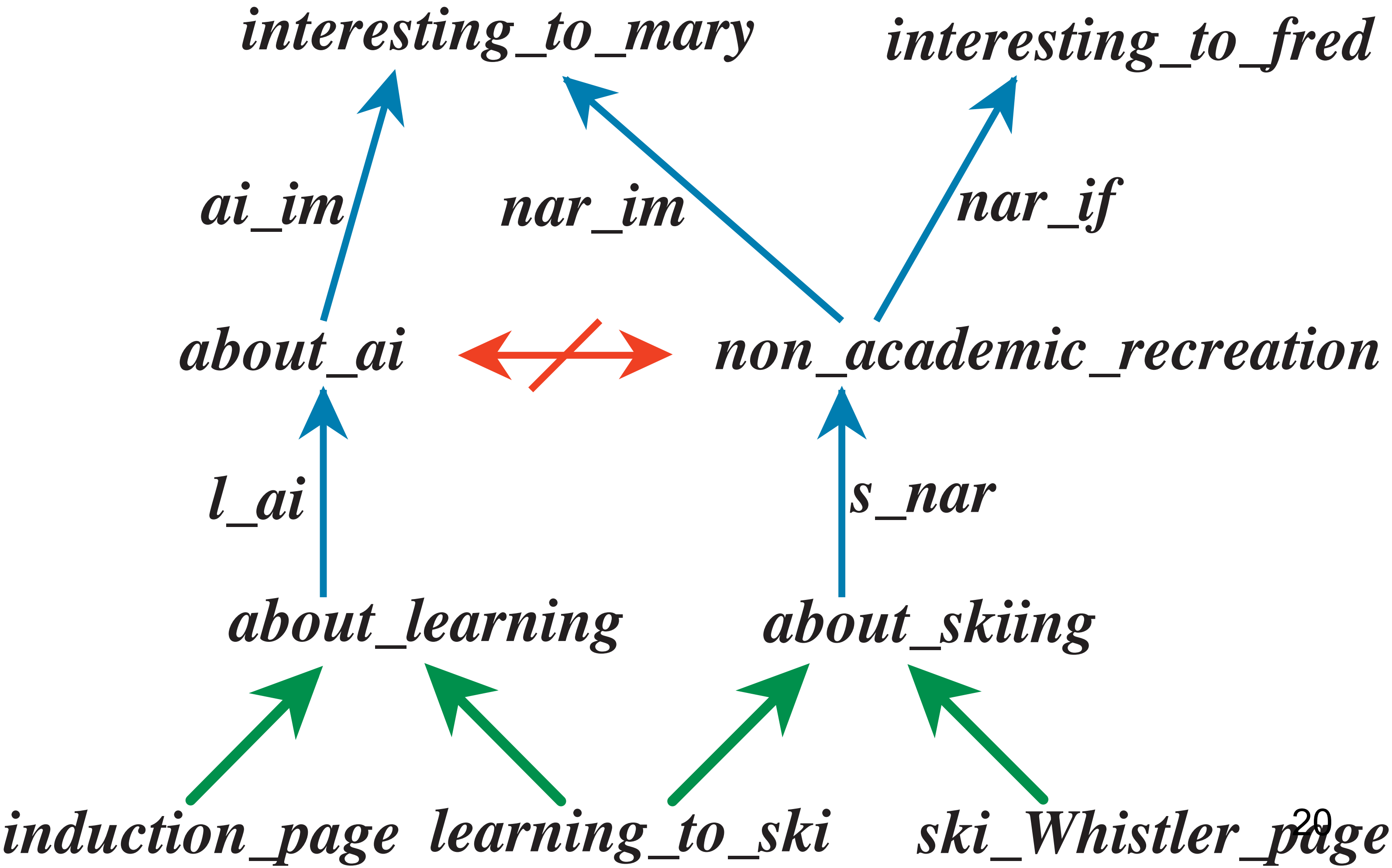
Diagram of the Default Example



Multiple Extension Problem

- What if incompatible goals can be explained and there are no cancellation rules applicable?
  What should we predict?
- For example: what if introductory questions are uninteresting, by default?
- This is the multiple extension problem.
- Recall: an extension of $\langle F, H \rangle$ is the set of logical consequences of F and a maximal scenario of $\langle F, H \rangle$.

Competing Arguments



Skeptical Default Prediction

- We predict g if g is in all extensions of $\langle F, H\rangle$.
- Suppose g isn't in extension E. As far as we are concerned E could be the correct view of the world. So we shouldn't predict g.
- If g is in all extensions, then no matter which extension turns out to be true, we still have g true.
- Thus g is predicted even if an adversary gets to select assumptions, as long as the adversary is forced to select something. You do not predict g if the adversary can pick assumptions from which g can't be explained.

Minimal Models Semantics for Prediction

Recall: logical consequence is defined as truth in all models.

We can define default prediction as truth in all minimal models.

Suppose M_1 and M_2 are models of the facts.

$M_1 <_H M_2$ if the hypotheses violated by M_1 are a strict subset of the hypotheses violated by M_2. That is:

$$\{h \in H' : h \text{ is false in } M_1\} \subset \{h \in H' : h \text{ is false in } M_2\}$$

where H' is the set of ground instances of elements of H.

Minimal Models and Minimal Entailment

- M is a minimal model of F with respect to H if M is a model of F and there is no model M_1 of F such that $M_1 <_H M$.

- g is minimally entailed from $\langle F, H \rangle$ if g is true in all minimal models of F with respect to H.

- Theorem: g is minimally entailed from $\langle F, H \rangle$ if and only if g is in all extensions of $\langle F, H \rangle$.

Abduction

Abduction is an assumption-based reasoning strategy where

- H is a set of assumptions about what could be happening in a system
- F axiomatizes how a system works
- g to be explained is an observation or a design goal

Example: in diagnosis of a physical system:
H contain possible faults and assumptions of normality,
F contains a model of how faults manifest themselves
g is conjunction of symptoms.

Abduction versus Default Reasoning

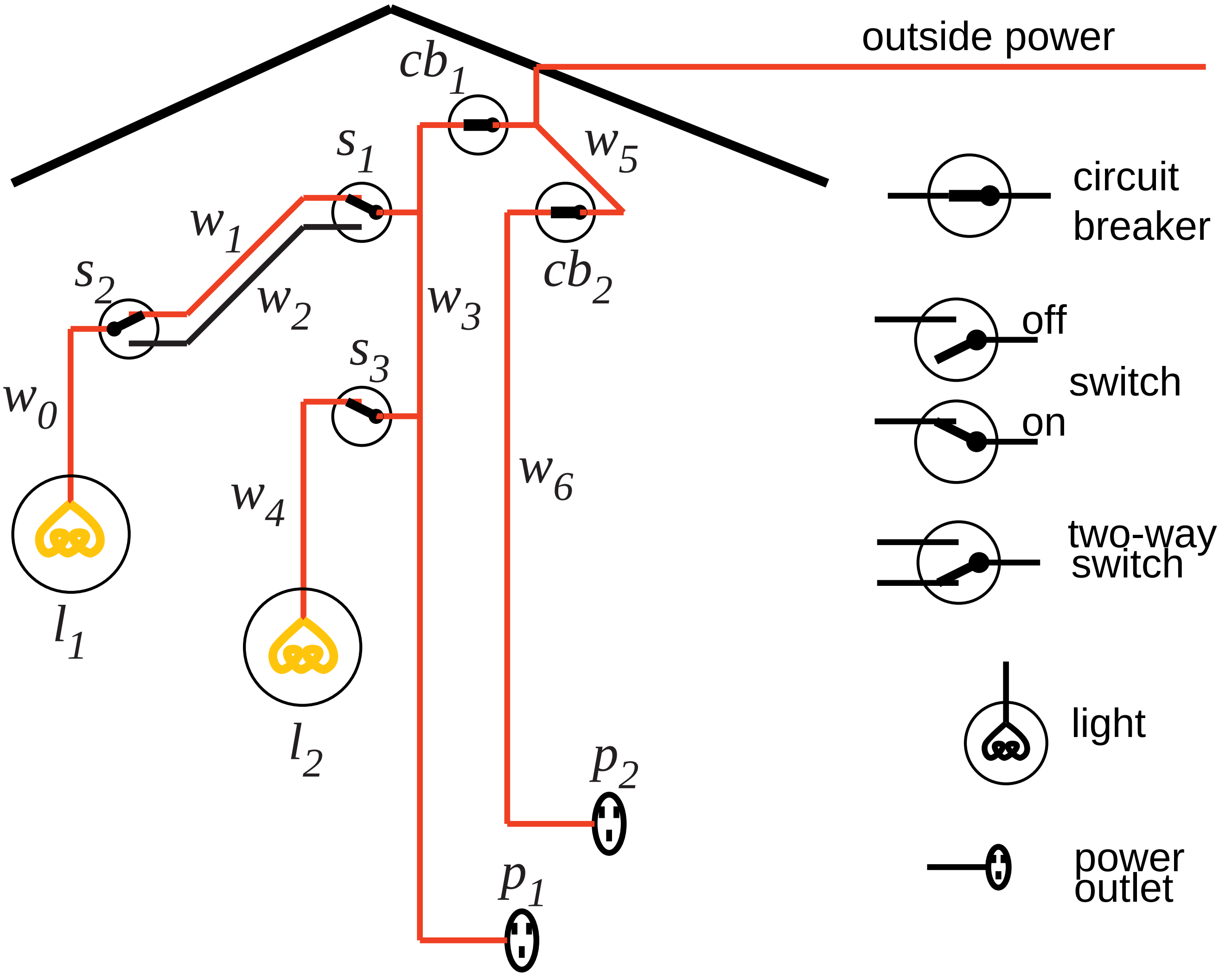
Abduction differs from default reasoning in that:

- The explanations are of interest, not just the conclusion.
- H contains assumptions of abnormality as well as assumptions of normality.
- We don't only explain normal outcomes. Often we want to explain why some abnormal observation occurred.
- We don't care if $\neg g$ can also been explained.

Abductive Diagnosis

- You need to axiomatize the effects of normal conditions and faults.
- We need to be able to explain all of the observations.
- Assumables are all of those hypotheses that require no further explanation.

Electrical Environment



$lit(L) \Leftarrow light(L) \;\&\; ok(L) \;\&\; live(L).$

$dark(L) \Leftarrow light(L) \;\&\; broken(L).$

$dark(L) \Leftarrow light(L) \;\&\; dead(L).$

$live(W) \Leftarrow connected\_to(W, W_1) \;\&\; live(W_1).$

$dead(W) \Leftarrow connected\_to(W, W_1) \;\&\; dead(W_1).$

$dead(W) \Leftarrow unconnected(W).$

$connected\_to(l_1, w_0) \Leftarrow true.$

$connected\_to(w_0, w_1) \Leftarrow up(s_2) \;\&\; ok(s_2).$

$unconnected(w_0) \Leftarrow broken(s_2).$

$unconnected(w_1) \Leftarrow broken(s_1).$

$unconnected(w_1) \Leftarrow down(s_1).$

$false \leftarrow ok(X) \wedge broken(X).$

assumable $ok(X)$, $broken(X)$, $up(X)$, $down(X)$.

Explaining Observations

- To explain $lit(l1)$ there are two explanations:
  $\{ok(l1), ok(s2), up(s2), ok(s1), up(s1), ok(cb1)\}$
  $\{ok(l1), ok(s2), down(s2), ok(s1), down(s1), ok(cb1)\}$
- To explain $lit(l2)$ there is one explanation:
  $\{ok(cb1), ok(s3), up(s3), ok(l2)\}$

Explaining Observations (cont)

- To explain $dark(l1)$ there are 8 explanations:

  $\{broken(l1)\}$

  $\{broken(cb1), ok(s1), up(s1), ok(s2), up(s2)\}$

  $\{broken(s1), ok(s2), up(s2)\}$

  $\{down(s1), ok(s2), up(s2)\}$

  $\{broken(cb1), ok(s1), down(s1), ok(s2), down(s2)\}$

  $\{up(s1), ok(s2), down(s2)\}$

  $\{broken(s1), ok(s2), down(s2)\}$

  $\{broken(s2)\}$

Explaining Observations (cont)

- To explain $dark(l1) \wedge lit(l2)$ there are explanations:

$\{ok(cb1), ok(s3), up(s3), ok(l2), broken(l1)\}$

$\{ok(cb1), ok(s3), up(s3), ok(l2), broken(s1), ok(s2), up(s2)\}$

$\{ok(cb1), ok(s3), up(s3), ok(l2), down(s1), ok(s2), up(s2)\}$

$\{ok(cb1), ok(s3), up(s3), ok(l2), up(s1), ok(s2), down(s2)\}$

$\{ok(cb1), ok(s3), up(s3), ok(l2), broken(s1), ok(s2), down(s2$

$\{ok(cb1), ok(s3), up(s3), ok(l2), broken(s2)\}$

Abduction for User Modeling

Suppose the infobot wants to determine what a user is interested in. We can hypothesize the interests of users:

$$H = \{interested\_in(Ag, Topic)\}.$$

Suppose the corresponding facts are:

$$selects(Ag, Art) \leftarrow$$
$$about(Art, Topic) \wedge$$
$$interested\_in(Ag, Topic).$$

$$about(art\_94, ai).$$

$$about(art\_94, info\_highway).$$

$$about(art\_34, ai). \qquad about(art\_34, skiing).$$

Explaining User's Actions

There are two minimal explanations of $selects(fred, art\_94)$:

$$\{interested\_in(fred, ai)\}.$$

$$\{interested\_in(fred, information\_highway)\}.$$

If we observe $selects(fred, art\_94) \wedge selects(fred, art\_34)$, there are two minimal explanations:

$$\{interested\_in(fred, ai)\}.$$

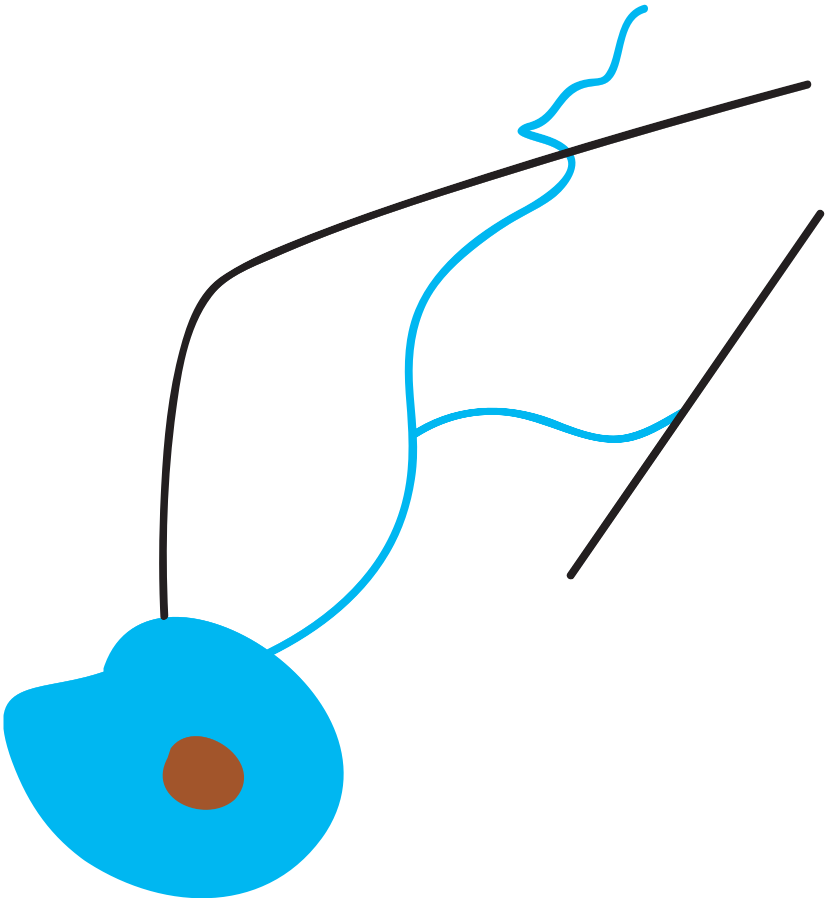
$$\{interested\_in(fred, information\_highway), \\ interested\_in(fred, skiing)\}.$$

Image interpretation

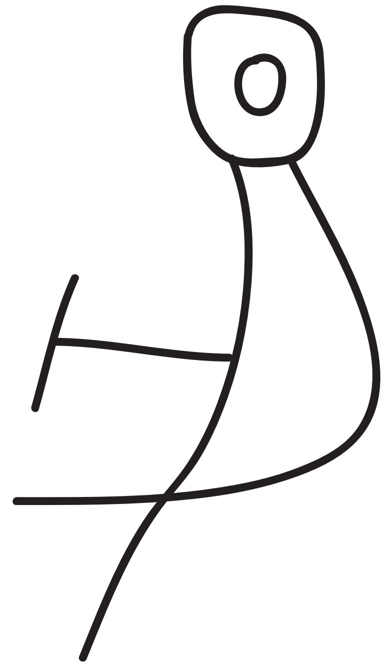
- A scene is the world that the agent is in.
- An image is what the agent sees.
- Vision: given an image try to determine the scene.
- Typically we know more about the *scene* $\rightarrow$ *image* mapping than the *image* $\rightarrow$ *scene* mapping.

Example Scene and Image

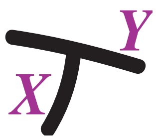
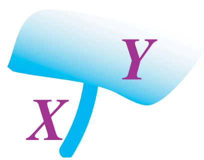
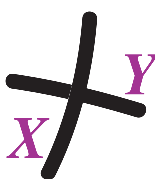
Scene



Image



Scene and Image Primitives

Scene Primitives	Image Primitives
land, water	region
river, road, shore	chain
joins(X, Y, E) 	tee
($E \in \{0, 1\}$ specifies which end of X)	
mouth(X, Y, E) 	
cross(X, Y) 	chi

Scene and image primitives (cont.)

Scene Primitives		Image Primitives
beside(C, R)		bounds(C,R)
source(C, E)		open(C,E)
loop(C)		closed(C)
inside(C, R)		interior(C,R)
outside(C, R)		exterior(C,R)

Axiomatizing the Scene $\rightarrow$ Image map

$chain(X) \leftarrow river(X) \vee road(X) \vee shore(X).$

$region(X) \leftarrow land(X) \vee water(X).$

$tee(X, Y, E) \leftarrow joins(X, Y, E) \vee mouth(X, Y, E).$

$chi(X, Y) \leftarrow cross(X, Y).$

$open(X, N) \leftarrow source(X, N).$

$closed(X) \leftarrow loop(X).$

$interior(X, Y) \leftarrow inside(X, Y).$

$exterior(X, Y) \leftarrow outside(X, Y).$

assumable $road(X), river(X), shore(X), land(X), \ldots$

assumable $joins(X, Y, E), cross(X, Y), mouth(L, R, E) \ldots$

Scene Constraints

$$false \leftarrow cross(X, Y) \wedge river(X) \wedge river(Y).$$

$$false \leftarrow cross(X, Y) \wedge (shore(X) \vee shore(Y)).$$

$$false \leftarrow mouth(R, L1, 1) \wedge river(R) \wedge mouth(R, L2, 0).$$

$$start(R, N) \leftarrow river(R) \wedge road(Y) \wedge joins(R, Y, N).$$

$$start(X, Y) \leftarrow source(X, Y).$$

$$false \leftarrow start(R, 1) \wedge river(R) \wedge start(R, 0).$$

$$false \leftarrow joins(R, L, N) \wedge river(R) \wedge (river(L) \vee shore(L)).$$

$$false \leftarrow mouth(X, Y, N) \wedge (road(X) \vee road(Y)).$$

$$false \leftarrow source(X, N) \wedge shore(X).$$

$$false \leftarrow joins(X, A, N) \wedge shore(X).$$

$$false \leftarrow loop(X) \wedge river(X).$$

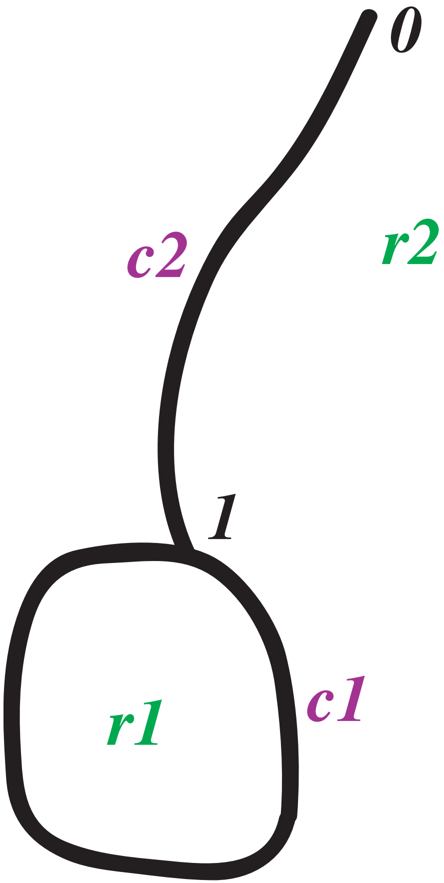
Scene constraints (continued)

$$false \leftarrow shore(X) \land inside(X, Y) \land outside(X, Z) \land$$
$$land(Y) \land land(Z).$$

$$false \leftarrow shore(X) \land inside(X, Y) \land outside(X, Z) \land$$
$$water(Z) \land water(Y).$$

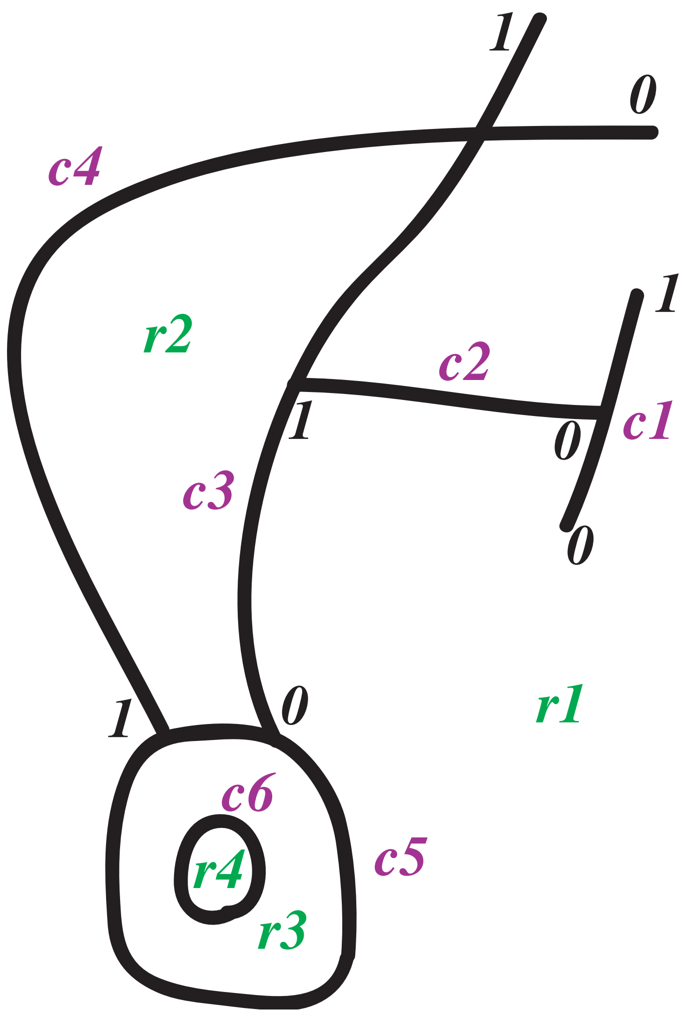
$$false \leftarrow water(Y) \land beside(X, Y) \land$$
$$(road(X) \lor river(X)).$$

Describing an image



$chain(c1) \wedge chain(c2) \wedge$
$region(r1) \wedge region(r2) \wedge$
$tee(c2, c1, 1) \wedge$
$bounds(c2, r2) \wedge$
$bounds(c1, r1) \wedge$
$bounds(c1, r2) \wedge$
$interior(c1, r1) \wedge$
$exterior(c1, r2) \wedge open(c2, 0)$
$\wedge\ closed(c1)$

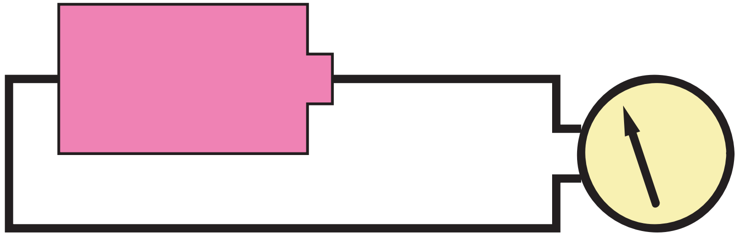
A more complicated image



$chain(c1) \wedge open(c1, 0) \wedge$
$open(c1, 1) \wedge region(r1) \wedge$
$bounds(c1, r1) \wedge chain(c2) \wedge$
$tee(c2, c1, 0) \wedge bounds(c2, r1)$
$\wedge\ chain(c3) \wedge bounds(c3, r1) \wedge$
$region(r2) \wedge bounds(c3, r2) \wedge$
$chain(c5) \wedge closed(c5) \wedge$
$bounds(c5, r2) \wedge$
$exterior(c5, r2) \wedge region(r3) \wedge$
$bounds(c5, r3) \wedge$
$interior(c5, r3) \wedge \ldots$

Parameterizing Assumables

Suppose we had a battery b connected to voltage meter:



To be able to explain a measurement of the battery voltage, we need to parameterize the assumables enough:

assumable $flat(B, V)$.

assumable $tester\_ok$.

$measured\_voltage(B, V) \leftarrow flat(B, V) \wedge tester\_ok$.

$false \leftarrow flat(B, V) \wedge V > 1.2$.

Evidential and Causal Reasoning

Much reasoning in AI can be seen as evidential reasoning, (observations to a theory) followed by causal reasoning (theory to predictions).

Diagnosis Given symptoms, evidential reasoning leads to hypotheses about diseases or faults, these lead via causal reasoning to predictions that can be tested.

Robotics Given perception, evidential reasoning can lead us to hypothesize what is in the world, that leads via causal reasoning to actions that can be executed.

Combining Evidential & Causal Reasoning

To combine evidential and causal reasoning, you can either

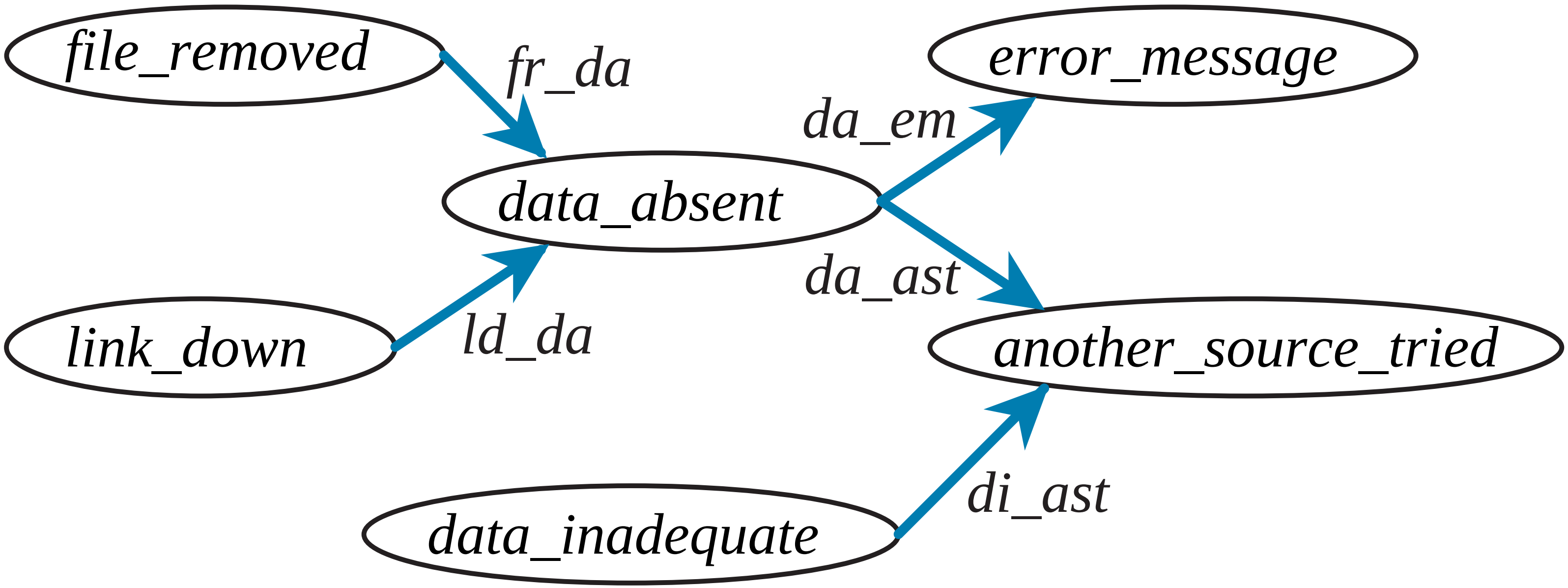
- Axiomatize from causes to their effects and
  - use abduction for evidential reasoning
  - use default reasoning for causal reasoning
- Axiomatize both
  - effects $\longrightarrow$ possible causes (for evidential reasoning)
  - causes $\longrightarrow$ effects (for causal reasoning)

  use a single reasoning mechanism, such as default reasoning.

Combining abduction and default reasonin

- Representation:
  - Axiomatize causally using rules.
  - Have normality assumptions (defaults) for prediction
  - other assumptions to explain observations
- Reasoning:
  - given an observation, use all assumptions to explain observation (find base causes)
  - use normality assumptions to predict from base causes explanations.

Causal Network



Why is the infobot trying another information source?

(Arrows are implications or defaults. Sources are assumable.)

Code for causal network

$error\_message \leftarrow data\_absent \wedge da\_em.$

$another\_source\_tried \leftarrow data\_absent \wedge da\_ast$

$another\_source\_tried \leftarrow data\_inadequate \wedge di\_ast.$

$data\_absent \leftarrow file\_removed \wedge fr\_da.$

$data\_absent \leftarrow link\_down \wedge ld\_da.$

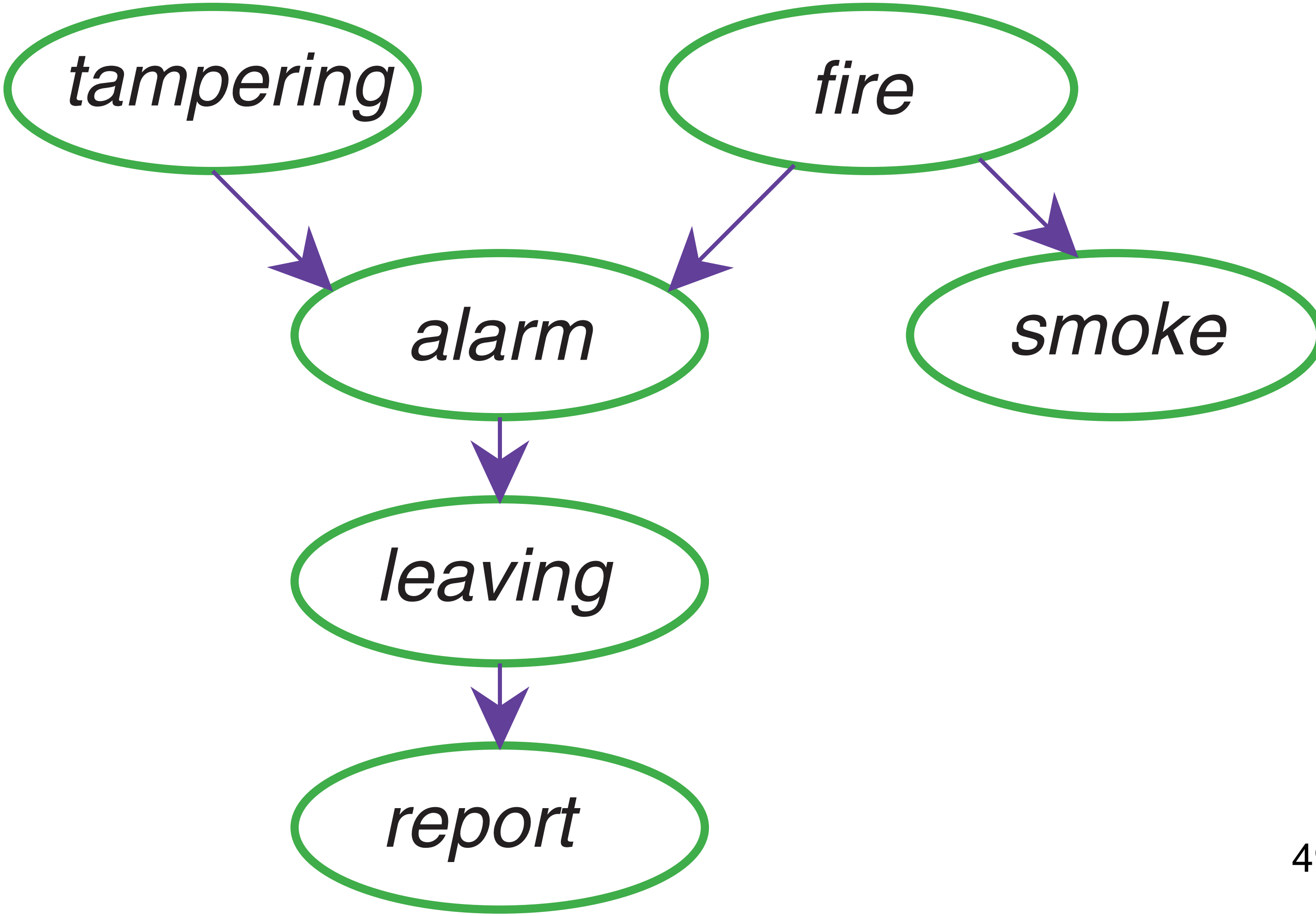
default $da\_em, da\_ast, di\_ast, fr\_da, ld\_da.$

assumable $file\_removed.$

assumable $link\_down.$

assumable $data\_inadequate.$

Example: fire alarm



Fire Alarm Code

$alarm \leftarrow tampering \land tampering\_caused\_alarm.$

default $tampering\_caused\_alarm.$

assumable $tampering.$

$alarm \leftarrow fire \land fire\_caused\_alarm.$

default $fire\_caused\_alarm.$

assumable $tampering.$

assumable $fire.$

$smoke \leftarrow fire \land fire\_caused\_smoke.$

default $fire\_caused\_smoke.$

Explaining Away

- If we observe *report* there are two minimal explanations:
  - one with *tampering*
  - one with *fire*
- If we observed just *smoke* there is one explanation (containing *fire*). This explanation makes no predictions about tampering.
- If we had observed *report* $\land$ *smoke*, there is one minimal explanation, (containing *fire*).
  - The smoke explains away the tampering. There is no need to hypothesise *tampering* to explain report.